analytical chemistry

NOTTINGHAM TRENT UNIVERSITY

Subscriber access provided by Nottingham Trent University

Article

Fluctuation Scaling, the Calibration of Dispersion, and the Detection of Differences

Rianne Holland, Roman Rebmann, Craig D. Williams, and Quentin S. Hanley

Anal. Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.analchem.7b02909 • Publication Date (Web): 11 Oct 2017

ACS Publications

Analytical Chemistry is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Fluctuation Scaling, the Calibration of Dispersion, and the Detection of Differences

Rianne Holland,[*] Roman Rebmann,[*] Craig Williams,[†] and Quentin S. Hanley[*]

[*]*School of Science and Technology*
Nottingham Trent University
Clifton Lane
Nottingham NG11 8NS
United Kingdom

[†]*School of Biology, Chemistry and Forensic Science*
University of Wolverhampton
Wulfruna Street
Wolverhampton
WV1 1LY
United Kingdom

Corresponding author: Q. S. Hanley (quentin.hanley@ntu.ac.uk)

Abstract: Fluctuation scaling describes the relationship between the mean and standard deviation of a set of measurements. An example is Horwitz scaling which has been reported from inter-laboratory studies. Horwitz and similar studies have reported simple exponential and segmented scaling laws with exponents (α) typically between 0.85 (Horwitz) and 1 when not operating near a detection limit. When approaching a detection limit the exponents change and approach an apparently Gaussian ($\alpha = 0$) model. This behavior is generally presented as a property of inter-laboratory studies which makes controlled replication to understand the behavior costly to perform. To assess the contribution of instrumentation to larger scale fluctuation scaling, we measured the behavior of two inductively coupled plasma atomic emission spectrometry (ICP-AES) systems, in two laboratories measuring thulium using 2 emission lines. The standard deviation universally increased with the uncalibrated signal indicating the system was heteroscedastic. The response from all lines and both instruments was consistent with a single exponential dispersion model having parameters $\alpha = 1.09$ and $\beta = 0.0035$. No evidence of Horwitz scaling was found and there was no evidence of Poisson noise limiting behavior. The "Gaussian" component was a consequence of background subtraction for all lines and both instruments. The observation of a simple exponential dispersion model in the data allows for the definition of a difference detection limit (DDL) with universal applicability to systems following known dispersion. The DDL is the minimum separation between two points along a dispersion model required to claim they are different according to a particular statistical test. The DDL scales transparently with the mean and works at any location in a response function.

Key words: ICP-AES, ICP-OES, fluctuation scaling, characteristic function, uncertainty functions, noise precision models, Horwitz curve, HorRat, power laws, Taylor's law, gain, mean variance, analytical methods, exponential dispersion models, dispersion calibration.

Introduction:

Fluctuation scaling presented in a variety of ways has been widely observed in chemical measurements.[1-12] Fluctuation scaling in chemistry is presented implicitly in some papers,[10] while in others it is described by a range of terms (noise precision models,[11] uncertainty functions,[3] characteristic functions,[2] Horwitz curve,[8] etc.). Here, it will be referred to as fluctuation scaling to provide a link into the extensive literature from other disciplines providing useful approaches to interpreting results.[13-22] While many of these articles are presented in terms of variance, standard deviation is used here to be consistent with the prior literature in analytical chemistry.

Following on from earlier work, Horwitz[7,8] noted a scaling behavior in inter-laboratory studies which was later recognized to follow the form: $s = 0.02c^{0.85}$. This is an example of a process following an exponential dispersion model of the form: $s = \beta \bar{x}^{\alpha}$. Chemical fluctuation scaling has been widely reported following this and related forms in the context of inter-laboratory comparisons; however, studies addressing how it arises are limited. Most discussions are empirical in nature[3,8] with few published investigations into the mechanisms producing the observed behavior. Hanley[12] recently reported that many basic measurements including balance readings, fluorescence intensity, and absorbance are inconsistent with either Horwitz behavior or the characteristic function. Much of chemical measurement follows simple or segmented exponential dispersion models with exponents that do no match either the Horwitz exponent ($\alpha = 0.85$) or the main exponents of the characteristic function ($\alpha = 1$; $\alpha = 0$). To our knowledge, no attempts have been made previously to trace the origins of observed scaling from raw measurements through to calibrated results. The experiments reported here were designed to evaluate instrument repeatability as assessed by the standard deviation of replicate measurements and not to assess accuracy.

Theory

The Horwitz curve belongs to the family of exponential dispersion models having the property:

$$s = \beta \bar{x}^{\alpha} \quad (1)$$

where $\bar{x}$ is the mean, s is the standard deviation, and α and β are process and dispersion constants, respectively. The dispersion parameter, β, is equivalent to s when $\bar{x}$ =1. Some presentations of this and related functions use population mean, μ, and population standard deviation, σ, rather than $\bar{x}$ and s. The nomenclature here is meant to clarify that empirical studies have finite n. In the limit as n approaches ∞, equation 1 approaches the population parameters. Additionally, in Horwitz and many related studies, $\bar{x}$ is considered to be concentration and given the symbol c. The general form given by equation 1 can be in any units.

Many models of the form of equation 1 are Tweedie models and the Horwitz curve is a special case of equation 1. It is possible that the Horwitz curve *may* conform to a Tweedie model from a Poisson-Gamma process.[12] The difficulty is that not all models of the form of equation 1 are Tweedie models and the Horwitz exponent arises from a wide range of complex processes producing inter-laboratory data which include dispersion from measurements (e.g. mass, volume, intensity, etc.) made at lower levels of abstraction. It is unknown to what extent expectations based on classical error propagation rules predict behavior at higher levels of abstraction.

In the context of this study, a number of characteristics of statistical distributions, dispersion models, and processes must be understood. Specifically, the distribution of replicates of a single sample may give apparently Normally distributed local behavior. This does not mean

the process is Normal or that it conforms to a Normal dispersion model. A Normally distributed appearance may mask a fundamentally different process. For example, Poisson and Gamma distributions can be approximated by a Gaussian distribution under certain conditions. Under these same conditions the Poisson and Gamma distributions will be good approximations to the Gaussian distribution. In a fluctuation scaling plot, a Gaussian model requires the exponent to be 0, which also implies the data are homoscedastic ($s = \beta\bar{x}^0$). A Poisson model requires the exponent to be 0.5 and a Gamma model has an exponent of 1.

The well-known Normal, Poisson, Compound Poisson, and Gamma distributions belong to the Tweedie class of exponential dispersion models.[14,15,21,22] For this reason Tweedie models have been of interest for modeling a wide range of systems. However, all models of the form of equation 1 are not Tweedie models. Tweedie models with exponents $0 < \alpha < 0.5$ do not exist[22] and analytical balances have been observed to fall in this region.[12]

Finally, the observation of a particular exponent does not prove a particular process. It can falsify a hypothesis and suggest alternatives but does not provide proof. As an example, if $\alpha = 0$ then the data set is homoscedastic, incompatible with Poisson or Gamma processes allowing them to be rejected, and *suggestive* of a Gaussian process. However, a host of non-Gaussian processes (e.g. adding uniformly distributed deviates to a set of numbers) can generate a fluctuation scaling plot that has $\alpha = 0$.

A range of alternative forms modeling observed fluctuation scaling behavior have been proposed including segmented forms.[2,11] These include noise precision models of various types[11] and the characteristic function advocated by Thompson which has the form:[2,3]

$$s = \sqrt{a^2 + (\beta\bar{x})^2} \qquad (2)$$

Similarly to the Horwitz curve, the characteristic function is usually presented in terms of σ rather than s and $\bar{x}$ given in concentration or mass fraction units. It is worth noting that versions of equation 2 were invoked well before Horwitz's classic paper,[1,6] and the general shape has been widely observed.[2,3,23] Thompson provided a semi-empirical derivation based on the observation that measurements typically include a portion where s is nearly constant ($s = a$) and another where s is proportional to signal strength or concentration ($s = \beta \bar{x}$). The standard rules for propagation of independent variances corresponding to these two regions produce equation 2.[2] As a general approach, both the Thompson and Horwitz presentations are limited. The Horwitz equation has been criticized for predicting that the standard deviation will go to zero as concentration goes to zero; a condition that is not usually observed.[2] There is evidence that the Horwitz exponent is not universal in inter-laboratory comparisons[3] and is not observed in a wide range of underlying measurements.[12] Similarly, the characteristic function has limitations. Many studies do not have a regime where $s = \beta\bar{x}$,[2,3] many measurements have segmented forms,[12] and others are not expected to on theoretical grounds (e.g. Poisson data[24]). Examples where $s \neq \beta\bar{x}$ include: analytical balances, fluorimeters, and spectrophotometers.[12] This leaves a range of open questions about how these shapes and observed behavior arise.

A shape similar to equation 2 can be obtained by subtracting a constant, κ, which is less than the minimum observed mean signal ($\bar{x}_{\min} > \kappa$) from $\bar{x}$ when presenting a data set following equation 1 (Equation 3 and Figure 1).

$$s = \beta\left(\bar{x} - \kappa\right)^{\alpha} \qquad (3)$$

If κ is close to $\bar{x}_{\min}$, equation 3 models background subtraction and creates a family of shapes similar to equation 2. The importance of equation 3 is that it allows the authentic background noise source ascribed to a in equation 2 *to be generated by the same process as the noise*

ascribed to the β term. Thompson assumed they were independent additive variances from different processes – a Gaussian noise floor and a region having constant relative standard deviation. The *a* and *β* terms can be from different independent process (see below equations, 4,5, and 6), but this is not a requirement to produce a similar shape. Equation 3 implies that an apparently Gaussian segment in a fluctuation scaling plot can arise from any model of the form of equation 1 for which has $\alpha \geq 0$ via any data processing adding or subtracting a constant. In the ICP data sets presented below, there is only evidence for a single exponential model of the form of equation 1 suggesting a single error generating process. The apparent Gaussian zone *under these circumstances* is a consequence of data processing not an authentic transition from one process to another.

Equation 3 also predicts the consequences of adding electronic or digital offsets at any stage in the measurement process by considering κ less than zero. This produces a zone where apparent α approaches ∞ (Figure 1). The important feature of this discussion is that if a single exponential dispersion model describes the error process then the background against which detection is assessed belongs to the same model and *s* is entirely predictable based on the dispersion model. Understanding these shapes allows a range of instrumental characteristics (auto-zero, background subtraction, electronic offsets, etc.) to be inferred when otherwise they would not be transparent.

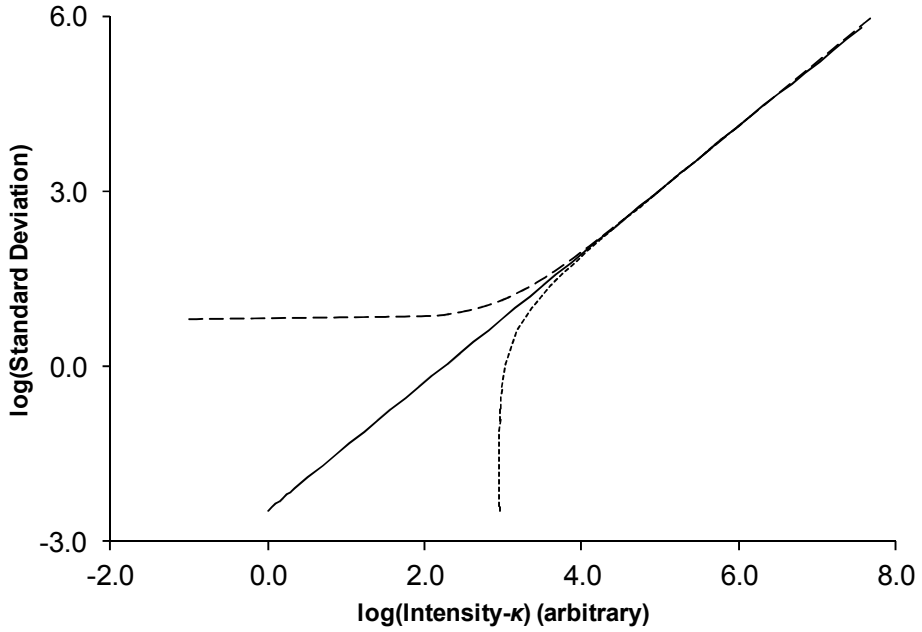


Figure 1: Simulations of an added offset (dotted line), subtracted background (long dashed line), and original model (solid line) by adjusting the value of κ in equation 3 while holding α and β constant.

More general models can be defined including summed, segmented, and mixed exponential dispersion models. Summed dispersion models are of the form of equation 4.

$$s = \sqrt{\sum_{i=1}^{n} \left(\beta_i \overline{x}^{\alpha_i} \right)^2} \qquad (4)$$

This form allows data sets exhibiting any exponent to be treated and, in the special case where $n = 2$, $\alpha_1 = 0$, and $\alpha_2 = 1$, this reduces to the Thompson characteristic function. This provides a form suitable for many circumstances. This type of model assumes there are n processes contributing to the error all of which are uncorrelated and contribute at every value of $\overline{x}$. An example of such a system is a CCD detector investigated over the range from read noise limited through Poisson noise limited behavior.[24]

Segmented exponential dispersion models are of the form of equation 5.

$$s = \begin{cases} \beta_1 \bar{x}^{\alpha_1} & for \quad (\bar{x} < t_1) \\ \beta_2 \bar{x}^{\alpha_2} & for \quad (t_1 \le \bar{x} < t_2) \\ \beta_2 \bar{x}^{\alpha_2} & for \quad (t_2 \le \bar{x} < t_3) \\ \quad \vdots & \\ \beta_n \bar{x}^{\alpha_n} & for \quad (t_{n-1} \le \bar{x} < t_n) \end{cases} \tag{5}$$

These models apply to systems where the process determining the error changes at threshold points. Thompson presented models with two[2] and three[4] segments but did not provide a general framework nor an explanation of how three segments might arise. Models of this type require a fundamental change in the process generating the data set at the threshold points. A published example of such a response is a fluorimeter assessed in a particular way.[12] Segmented behavior would also be expected when a measurement system saturates. As the signal approaches the threshold for saturation, the standard deviation will decrease.

Mixed exponential dispersion models are hybrids of summed and segmented forms. An example is a CCD imaging detector having a fixed readout noise.[25,26] At low signal, readout noise dominates. As the device is exposed to increasing amounts of light, it follows Poisson statistics. When it approaches saturation, depending on system characteristics it might return to a fixed readout noise. Such a system is described by equation 6.

$$s = \begin{cases} \sqrt{\left(\beta_1 \bar{x}^{\alpha_1}\right)^2 + \left(\beta_2 \bar{x}^{\alpha_2}\right)^2} & for \quad (\bar{x} < t_1) \\ \beta_3 \bar{x}^{\alpha_3} & for \quad (t_1 \le \bar{x} < t_2) \end{cases} \tag{6}$$

It should be noted that there is no requirement for measurements to follow models following 1, 4, 5, and 6. For example, spectrophotometers measuring absorbance follow more complex models;[12] however, the measured light intensities used to compute the absorbance will in many cases conform to these models.

When a measurement follows the form of equation 1 either generally or locally as part of more complex models, the gain and/or relative gain can be defined.[12] The gain, *G*, is a

constant multiplier applied at any point during the measurement and calibration process. The effects of gain can be predicted by equation 7.

$$s = G^{(1-\alpha)}\beta\,\bar{x}^{\alpha} \tag{7}$$

An example of a measurement with variable gain is an electron multiplying CCD measuring light.[27] Relative gain, G_R, is a way to define the relative behavior between two systems believed to conform to the same scaling law.[12,28] For example, if two instruments indexed 1 and 2 have the same α and β, then

$$G_R = \frac{G_1}{G_2} = R^{\frac{1}{(1-\alpha)}} \tag{8}$$

where R is the ratio of measured pre-exponential factors,

$$R = \frac{G_1^{(1-\alpha)}\beta}{G_2^{(1-\alpha)}\beta} = \left(\frac{G_1}{G_2}\right)^{(1-\alpha)} \tag{9}$$

These considerations are useful in the context of a simple calibration procedure consisting of multiplication of a measured data point by a constant. Simple calibration of this type is equivalent to application of a gain factor and can be introduced at any stage in the measurement and calibration process. A special case exists when $\alpha = 1$ resulting in $R = 1$ and $G_R = 1$ for all values of G_1 and G_2.

Equations 3 and 7 allow a general treatment of the impact of either classical ($\bar{x} = mc + b$) or inverse ($c = m\bar{x} + b$) linear calibration on a dispersion relationship of the form of equation 1. Here, m and b are the slope and intercept of the calibration and c is the concentration. With suitable definitions for G and κ, the predicted models become:

$$s = G^{(1-\alpha)}\beta\left(\bar{c} + \kappa\right)^{\alpha} \tag{10}$$

This is very similar in shape to Thompson's characteristic function when α = 1. The key difference is that the squared version includes cross terms while Thompsons's does not making it closer to the form of root quadratic NPMs.[11]

Generalized Difference Detection.

These considerations allow the definition of a difference detection limit (DDL) applicable to all data sets conforming to a known dispersion model relating *s* to $\bar{x}$. We define the DDL as the minimum distance between two mean signals required to conclude they are different to a specified degree of confidence. This can be used to estimate whether a difference will be detectable using a range of statistical criteria. Measurements following equation 1 are easiest to apply, however, systems conforming to equations 4, 5, 6 or more complex forms, though less convenient, can also be used to estimate a DDL.

As an example, suppose we wish to detect a difference for a process following an exponential dispersion model of the form of equation 1 using an *F*-ratio. This requires an assumption that at each position along the model the local distribution is approximately normal. With this assumption, a general formula is obtained, $\left(\bar{x}_2 / \bar{x}_1\right)^{\alpha} = \sqrt{F_{crit}}$. This one compact equation defines the requirements for *F*-test difference detection for every model meeting the assumptions of approximate normality and following the form of equation 1. It applies to every $\bar{x}$ and the only further requirement is to adjust F_{crit} to match the decision requirement for the DDL. If $\bar{x}_1$ = 1, α = 1, and *n* = 10 and 99% confidence (F_{crit} = 5.35) is needed, then $\bar{x}_2$ = 2.31 and the DDL ($\Delta\bar{x}_{DDL} = \bar{x}_2 - \bar{x}_1$) will be 1.31. A mean value of 2.31 represents the detection of a difference from 1 at 99% confidence. The DDL scales with $\bar{x}_1$ (e.g. for α = 1, and *n* = 10 when $\bar{x}_1$ = 10, the 99% confident DDL will be 13.1). This example has been selected as it represents a somewhat unusual variant of detection limit methodology by

comparing variance using the *F*-ratio rather than comparing means. The *F*-ratio case is interesting because illustrates how to use the process parameter, α. A DDL cannot be defined using an F-ratio for a system with a Gaussian dispersion model ($\alpha = 0$) due to the DDL going to infinity as α approaches 0. However, as α gets large the DDL becomes vanishingly small.

DDLs can be defined using the more conventional approach of comparing means and if needed the statistical criteria adjusted to minimize false negatives. The definition has generality while avoiding a range of issues including: background subtraction, translating raw data units into concentration units below a limit of quantification, and being restricted to a specific part of a response function. Although it has been illustrated here with an *F*-ratio giving a convenient general formula, a host of other statistical methods could be used.

Experimental Section

ICP-OES

Thulium solutions (0.07 ppb to 10,000 ppm) were prepared from the chloride ($TmCl_3 \cdot 6\ H_2O$ (Aldrich, UK)) in dilute nitric acid. This metal was selected to minimize the possibility of contamination effects at low concentration. ICP measurements were made using two commercial ICP-AES systems (Optima 2100D, Perkin Elmer; and ICP-OES 5100, Agilent) equipped with auto samplers using the parameters indicated in Table 1. Within each trial a set of 10 replicate measurements were made to assess the mean and standard deviation.

Parameter	Optima 2100D	Agilent 5100
Gas flow rate	15 L/min	12 L/min
Auxiliary gas flow	0.2 L/min	1.0 L/min
Nebulizer flow rate	0.8 L/min	0.7 L/min
RF power	1300 Watts	1300 Watts
Pump flow rate	1.50 mL/min	1.50 mL/min

Tm Wavelengths	313.126 nm 346.220 nm	313.126 nm 346.220 nm

Table 1: ICP-AES parameters.

The descriptive statistics and scaling relationships were fit using log transformed data and parameters were estimated by standard regression methods in MS Excel with the Real Statistics Resource Pack (http://www.real-statistics.com/) except for the segmented and bent cable fits to the data which were done using R (Version 3.31) with the Segmented (Version 0.5-2.1) and SiZeR (Version 0.1-4) packages, respectively. The Anscombe-Glynn[29] test for kurtosis was done using the Moments package (Version 0.14). Iterative fitting with the Solver add-in of MS Excel was used to determine the offsets in instrument B by maximizing the R^2 of the log-log linear regression while adjusting the value of the offsets.

Results and discussion

Uncalibrated Instrument Response

The raw uncalibrated instrument fluctuation scaling response for the two instruments (Figure 2) to Tm gave two distinct responses. Instrument A (Figure 2a) gave a single exponential dispersion model ($s = 0.00316\bar{x}^{1.086}$) independent of the wavelength of observation. There was minimal sign of saturation (although the highest sample at 313 nm did not return a result) and no observable electronic offset. Instrument B (Figure 2b) showed three regions in each response suggesting the form of equation 5. At the lowest concentrations of Tm, a near vertical response was observed. This was attributed to an arbitrary offset in the electronics of the instrument (equation 3). At intermediate concentrations, behavior similar to that seen in instrument A was observed. Although the slopes for the intermediate range appear different in the figure (Figure 2b), the 95% confidence intervals overlapped indicating only a single dispersion model was needed. At the highest concentrations saturation occurred. Segmented

analysis (the fitted lines in Figure 2b) indicated the saturation breakpoints corresponded to values above 300 ppm Tm.

To allow better comparison between the instruments, the data above the second breakpoint from Instrument B was removed and offsets fit to the data by optimizing the fit to a power law (Figure 2c) using equation 3 to estimate the constant. This left a simple dispersion model of the form of equation 1 ($s = 0.00345\bar{x}^{1.093}$) for Instrument B giving remarkable agreement between the two instruments (Figure 2d). After offset adjustment, there were no significant differences in dispersion or process parameters between the two instruments excepting the saturation region above 300 ppm in instrument B. The observed α value near one indicates that gain has little or no effect on a data set. Specifically, equation 7 reduces to $s = G^{(0.093)}\beta\bar{X}^{1.093}$ for this data implying that s will remain within ±30% of the original values for values of G over the range 0.1 to 10. This characteristic of exponential dispersion models following equation 1 with α near one probably gives rise to the very close correspondence between the two instruments.

The data are conclusive on a number of points. Homoscedasticity can be firmly rejected and has no place in any discussion of ICP-AES. Homoscedastic data requires $\alpha = 0$ which is not observed. Poisson noise limitation can also be rejected for these instruments and conditions. For Poisson noise, α would be expected to be 0.5 rather than 1.09. The exponent is above 1 at > 95% confidence indicating a modest super-linear relationship between mean and standard deviation. Relative signal to noise (repeatability) will be improved at lower intensities and the exponent, α, provides a framework for deciding this. In general, when $\alpha < 1$ increasing the signal level by increasing the sample concentration will be beneficial; when $\alpha = 1$ the quality of the measurement will be independent of signal level and largely independent of concentration; and when $\alpha > 1$ the relative uncertainty is improved by lower signals. There

was no Gaussian noise segment. Before removing the offset from Instrument B, the opposite is seen; a near vertical segment. If there is a Gaussian noise floor, it is impossible to view under useful measurement conditions due to the continuum background from the plasma. A single definable dispersion parameter (β = 0.0035) describes both instruments well. This indicates a percent relative standard deviation (PRSD) of 0.35% when $\bar{x}$ = 1 which increases to ~1% at the top of the range. These dispersion (β = 0.0035) and process (α = 1.09) parameters may be a characteristic of modern ICP-instruments and claims of improvement should be referenced against them rather than detection limits. Detection limits are controversial due to lack of modernization by some bodies.[30] However, using the formula presented earlier for an F-ratio defined DDL, a difference between two mean values can be claimed whenever $\left(\bar{x}_2 / \bar{x}_1\right)^{1.09} \geq F_{crit}$. Under the definitions of the DDL, it applies anywhere along the curve shown in figure 2d and will scale appropriately with signal strength.

A fluctuation scaling law with α near 1 can be generated by sample introduction variability. In modern ICP, this might be caused by sample pump fluctuations or other stochastic aspects of nebulizers, spray chambers and spray entrainment. Any such fluctuations would affect all samples independent of concentration. Noisy sample introduction can also explain the asymptotic behavior of Thompson's characteristic function and any study where α is near 1. Here, there is no sign the exponent changes when approaching the plasma background. The plasma background of the high intensity line (346.220 nm) is coincident with an analytically useful portion of the fluctuation curve for the less intense line (313.126 nm). Noisy sample introduction cannot explain this unless the fluctuations are the result of the plasma interacting with the aqueous sample matrix.

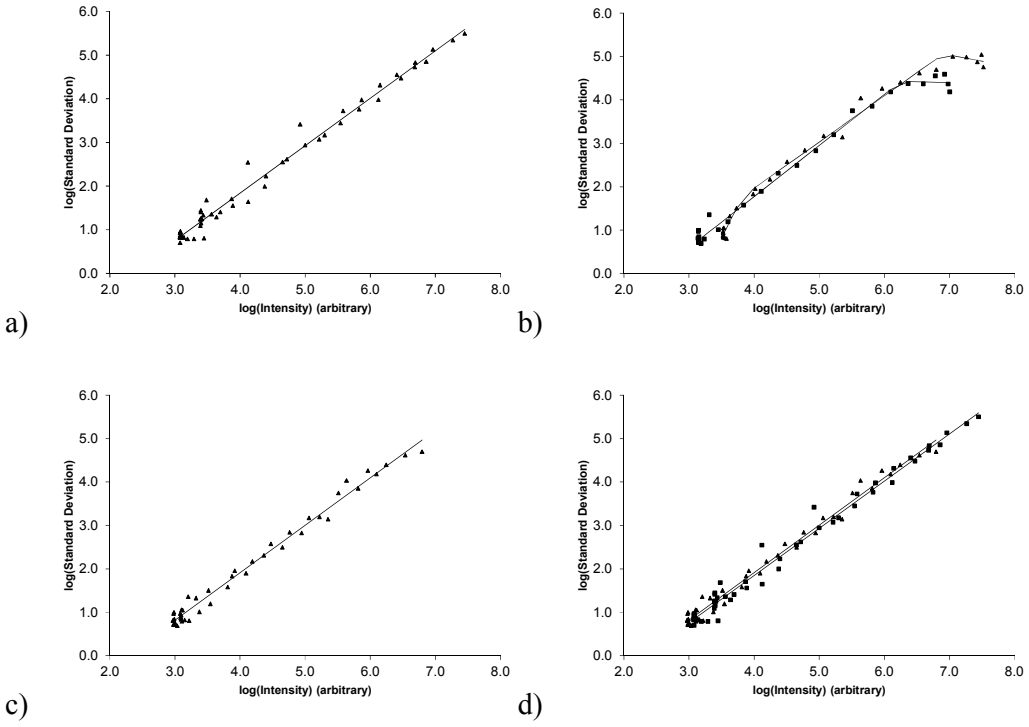


Figure 2: Raw uncalibrated instrument response to Tm at 313.126 nm and 346.220 nm. The two lines gave identical results for Instrument A (panel a) while Instrument B exhibited signs of offsets and saturation (panel b). In panel b, the lines represent the best fit piecewise model with two breakpoints the squares corresponding to 313.126 nm while the triangles correspond to 346.220 nm. After adjusting for the offset and the saturation (panel c) a single fluctuation scaling law described the response of both instruments (panel d). For comparison best fit lines are provided for Instrument A (squares) and Instrument B (triangles).

Higher Moments (Skew and Kurtosis) of the measured Responses

Higher moments were investigated to infer information about underlying error generating processes. The first approach was to construct a population skew-kurtosis plot using each sample's ten replicates to compute skew and kurtosis (Figure 3). The second approach was to construct a normalized residual histogram for instrument A by subtracting the sample mean from each of the 10 replicate measurements and dividing by the computed standard deviation from the scaling law (Figure 4). The behavior of the Tweedie models for $\alpha = \{0, 0.5, 1\}$ which correspond to Normal, Poisson, and Gamma distributions, respectively, were considered as reference conditions for the data (Figure 3). The normal distribution has *skew* = 0 and *kurtosis* = 3. The Poisson distribution is parameterized by the mean, which also defines the skew and

kurtosis such that $kurtosis = skew^2 + 3$. The gamma distribution has shape, k ($k > 0$), and scale, θ ($\theta > 0$), parameters. The shape parameter defines the skew ($skew = 2/\sqrt{k}$) and kurtosis ($kurtosis = (6/k) + 3$) leading to $kurtosis = 1.5 skew^2 + 3$. Skew and Kurtosis are otherwise limited depending on n (c.f. Cox's tutorial and review[31]).

$$|skew| \leq \frac{n-2}{\sqrt{n-1}}; kurtosis \leq \frac{n^2 - 3n + 3}{n-1}$$

In the current study (n = 10), these limits were 8/3 and 73/9 for skew and kurtosis, respectively. Finally, a lower bound on the relationship between these two moments exists, $kurtosis \geq skew^2 + 1$.[32,33] As can be seen (Figure 3), the data lie within these limiting regions. Although similar plots to Figure 3 have been used to characterize scaling behavior in car paints,[34] they were unable to definitively show obvious correspondence to normal, Poisson, or gamma distributions.

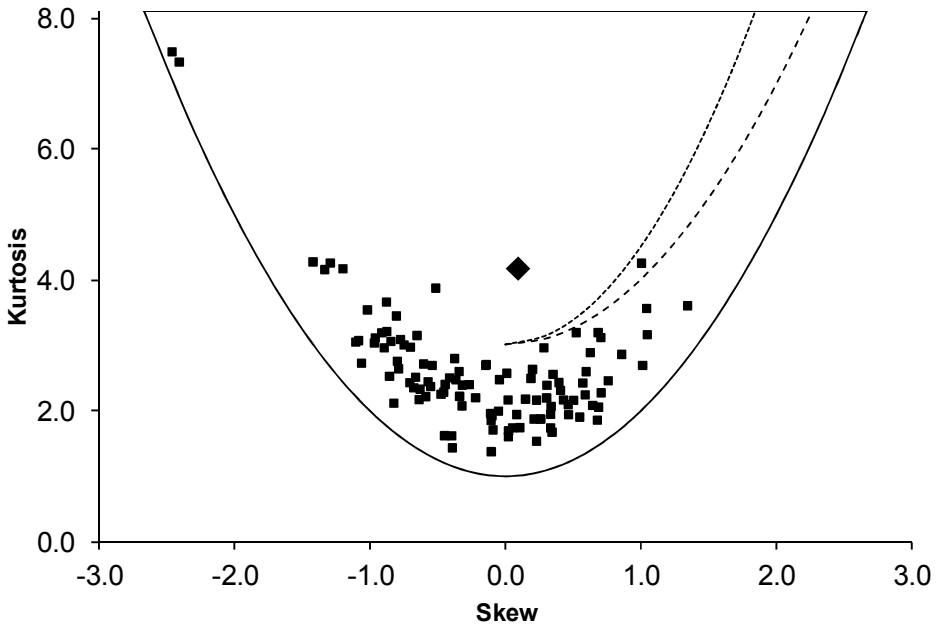


Figure 3: Skew and kurtosis plot of the raw uncalibrated instrument responses to Tm at 313.126 nm and 346.220 nm. The solid lines represent the theoretical bounds for n = 10. The upper short dashed line corresponds to ideal gamma distributed data and the middle long dashed line represents the behavior of Poisson distributed data. A normal distribution would

appear at (3,0). The position of the normalized residuals (n = 528) shown as a histogram in figure 4 is given by the large diamond.

The residual histogram (Figure 4) and analysis of the residuals directly were of greater assistance. A variety of distributions were investigated to fit the histogram, with none showing clear superiority. Logistic and Cauchy-Lorenz distributions fit better than the normal distribution, but not significantly so. The data showed significantly ($p < 0.0002$) greater kurtosis than expected from a normal distribution as assessed by the Anscombe-Glynn test.[29] The analysis also revealed 2 probable outliers in the data set. These observations indicate that currently, there is limited understanding of the types of error generation processes defining models of chemical fluctuation scaling and the resultant local and general distributions applicable to this basic chemical measurement. Future work will provide an opportunity to carry out more wide ranging hypothesis testing to identify the underlying processes.

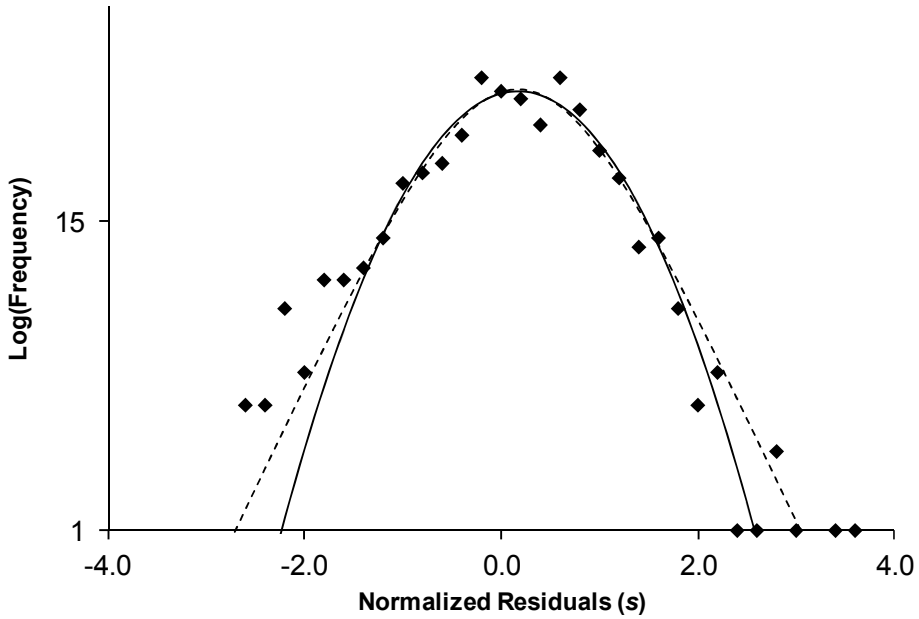


Figure 4: Normalized residual histogram for all measurements from Instrument A. Best fit scaled normal (solid line) and logistic (dashed line) distributions are shown on a log axis to highlight correspondence in the tails of the observed data. Initially, the residuals (n = 530) had skew and kurtosis of -2.25 and 22.49 due to two measurements which appeared to be outliers.

These two values also caused the two points in the extreme upper left of Figure 3. These points had little influence on the histogram but strongly affected the computed higher moments. After removal skew and kurtosis became 0.09 and 4.17, respectively.

Uncalibrated Instrument Response with Background Subtraction

All four data sets (two instruments measuring Tm at two wavelengths) coincide in the fluctuation scaling plot following offset adjustment. As a result, background signal sometimes coincide with an analytically useful signal level for another line or instrument. In ICP-AES, there is a continuum background present that varies with wavelength. Subsequent subtraction (Figure 5) produces an apparently Gaussian zone approximating the form of equation 2. This feature was not due to a Gaussian process or a separate independent noise source as is commonly supposed. Rather, the shape can be created by subtracting a constant (the background) from the means when constructing a fluctuation scaling plot.

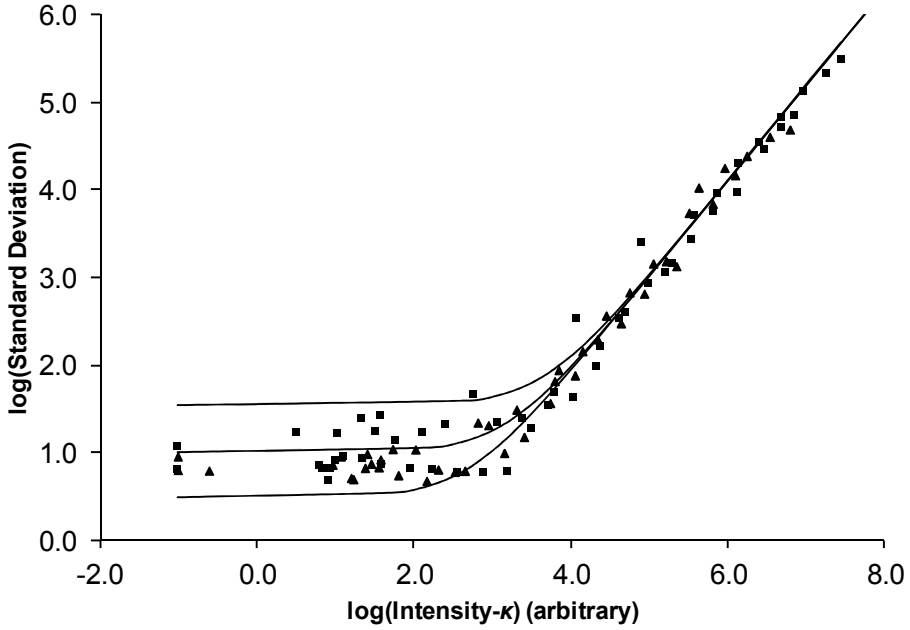
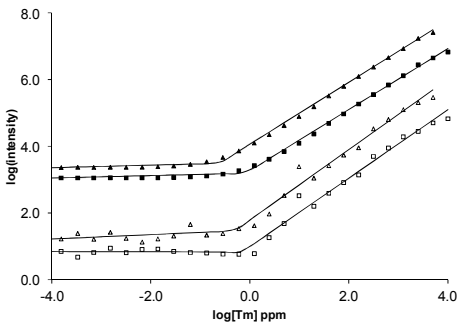


Figure 5: Raw uncalibrated instrument response to Tm at 313.126 nm and 346.220 nm following background subtraction for Instrument A (squares) and Instrument B (triangles). The subtracted background values were 1192 and 2466 for Instrument A and 1382 and 3323 for Instrument B. The lines are obtained by setting κ = {500, 1500, 4500} in the model $s = 0.0035\left(\overline{x} - \kappa\right)^{1.09}$ to bracket the observed data.

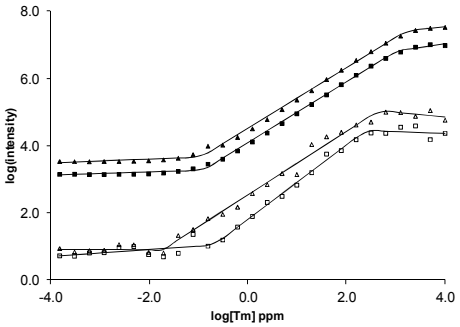
Calibration of the instrument response.

Calibration computations influence the fluctuation scaling behavior in ICP-AES (Figure 6). The simplest calibration would consist of multiplication by a constant. This is a simple gain adjustment and for exponential dispersion models with exponents near 1 has little or no impact on the fluctuation scaling behavior (e.g. equation 7). For the wide range of [Tm] investigated, such a simple procedure was insufficient. In order to understand the impact of calibration on the behavior of the dispersion model, the mean and standard deviation of measured intensity were plotted as a function of [Tm] (Figure 6a and 6b). Attempts to fit the data were made using segmented (Figure 6a and 6b) and bent cable regression (Figure 6c and 6d). Neither was particularly satisfactory as indicated by residual plots (Figure 6e and 6f). Bent cable regression was better, but in neither case were the residuals randomly distributed. Segmented and bent cable regression methods both assume the underlying behavior is linear and differ only in that bent cable regression tests whether the transition between segments is smooth. The goal of the exercise was to attempt a conversion from raw data to concentration units while having minimal impact on the dispersion model, particularly near the upper and lower extremes of the data set. This is remarkably difficult and there is limited discussion of the overall shape of the raw response function in ICP-AES in the literature, particularly in the case of modern instruments. Studies over wider ranges have tended to use *ad hoc* segmented calibrations over relatively restricted ranges. Log-log fits after background subtraction gave good approximations to the response over a restricted range (figures 6g and 6h); however, this procedure will produce biased estimates of the concentration at the extreme ends. Instrument A fit well to a straight line; however, Instrument B exhibited a slight curvature. Application of these calibrations to the restricted ranges identified in figures 6g and 6h yielded identical predicted vs. actual [Tm] responses for both instruments using both emission lines (figure 6i). The fluctuation scaling in calibrated units (figure 6j) exhibited an upward trend in some cases

in the foot of the fluctuation scaling curves. This is not considered significant, as there is some upward bias introduced by the logarithmic processing after background subtraction. Over the range from 1-300 ppm there was no difference between the instruments when measuring Tm using either emission line. The position of the foot of the scaling function varied with Instrument B consistently lower than Instrument A, an observation consistent with the measured resolution of the two instruments. Instrument B had an FWHM of 0.02 nm for the Tm 346.220 nm line while the FWHM for Instrument A was 0.03 nm.



a)



b)

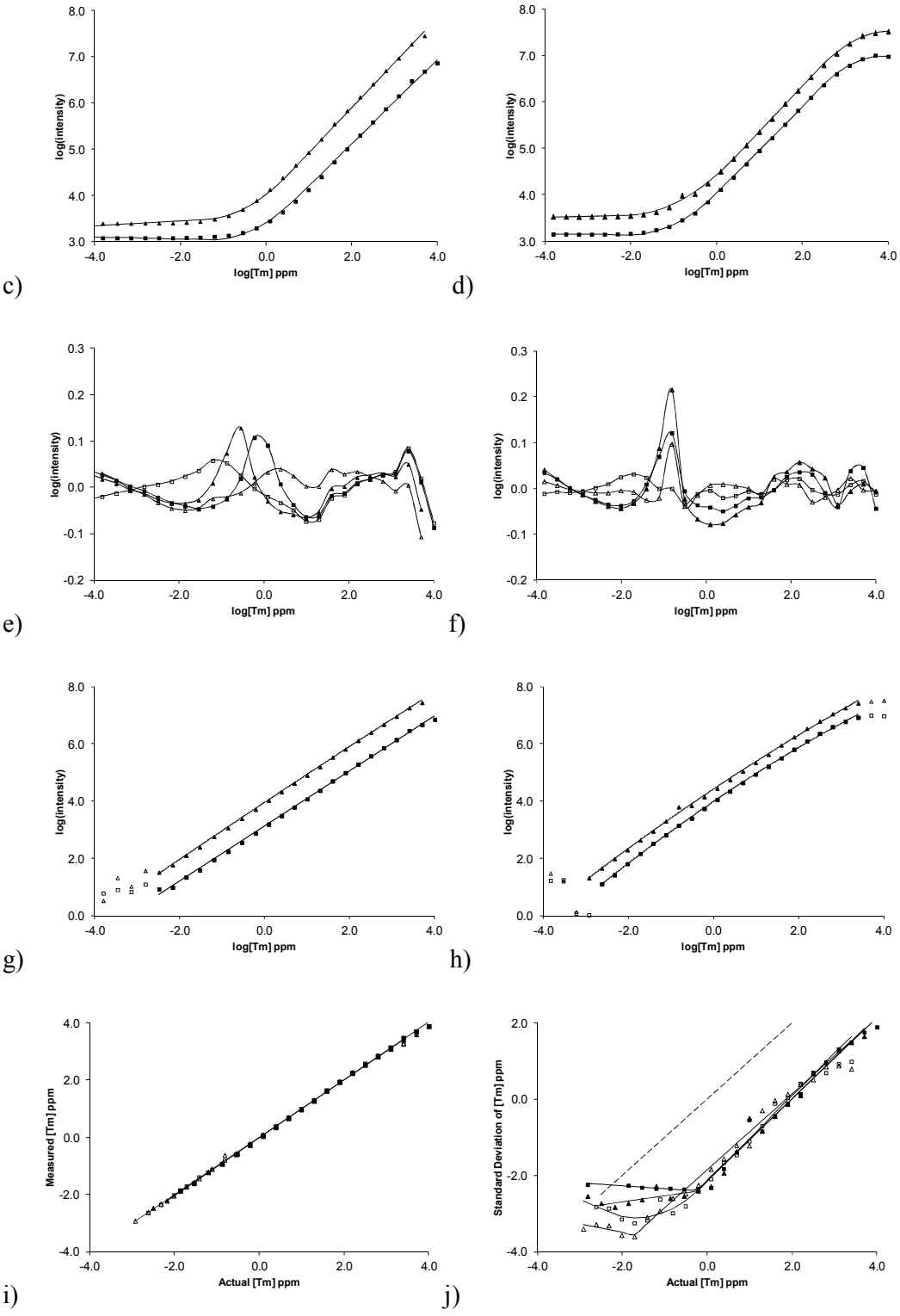


Figure 6: Raw and calibrated responses of Instrument A (a, c, e, and g) and B (b, d, f, and h). The raw responses to Tm concentrations over 8 orders of magnitude (a and b) showing average intensity (filled symbols) and the standard deviations (open symbols) for the 346.220 nm (triangles) and 313.126 nm (squares). Lines in panels (a) and (b) correspond to segmented regression models for these indicators. Bent cable regression models for the intensity data (c and d) gave better fits as indicated by the residual plots (e and f). In the residual plots, the

filled symbols correspond to the segmented regression and the open symbols to the bent cable regression. Background subtraction followed by linear (g) or polynomial regression (h) yielded a useful range of approximately 6 orders of magnitude. The location of the line indicates the region over which the regression was valid. Outside this line, some individual data points fell below 0 after background subtraction. The calibrated response (i) covered a wide range with all lines in all instruments producing nearly identical data. The calibrated fluctuation scaling (j) exhibited nearly identical behavior between 1-300 ppm, however some variation was observed elsewhere. The dashed line corresponds to the results of calibrated response (h).

Conclusion

This fluctuation scaling study demonstrated that the uncalibrated data from ICP-AES systems consisting of the measured intensity of light of specific wavelengths emitted from a plasma is consistent with an exponential dispersion model. Excepting a region of saturation above 300 ppm in one instrument, the model applied universally to all the lines monitored by two instruments manufactured approximately 10 years apart which were operated in different laboratories. The data were universally heteroscedastic and exhibited no evidence of Gaussian or Poisson models over any part of the response. Outside the saturated region in the one instrument, there was no evidence of summed, segmented, or mixed dispersion models as represented by equations 2, 4, 5, and 6. There was evidence of an electronic offset in one of the instruments consistent with equation 3.

The raw data fluctuation scaling law exponent suggested the Tweedie models conforming to gamma ($\alpha = 1$) or positive stable ($1 < \alpha < 1.5$) distributions might apply. Neither was particularly satisfying and further evidence for these was not found nor have they been proposed previously as important for this measurement. Fluctuation scaling exponents near 1 can be readily explained by variable sample introduction. For this to explain the current data set, the aqueous sample matrix would have to interact with the plasma to create fluctuations because background signals sometimes coincided with those from detectable Tm along the

scaling law. This indicates a large gap in our current understanding of appropriate statistical distributions to model chemical measurements and a robust explanation of the Horwitz exponent remains elusive. It also highlights a great opportunity to improve routine data processing methods beyond the homoscedastic assumptions of classical linear regression and related procedures. The value of the scaling law exponent ($\alpha = 1.09$) indicates that the method is nearly gain independent (see equation 7) and within some limits nearly concentration independent. This characteristic gives robustness to the scaling law suggesting it will be a widely reproducible benchmark.

Subtraction of a constant from the mean prior to computing the exponential fluctuation scaling model produced an apparently Gaussian zone ($\alpha \sim 0$) followed by linear scaling consistent with many previous reports.[1-3,30] Here, the shape is created by background subtraction which yields a form similar to a root quadratic noise precision model.[11] Such a shape should not be interpreted as having separate noise generation processes as expected from equations 2, 4, 5, and 6. This interpretation is clearly incorrect for ICP-AES and is likely to be incorrect in many others. Any apparently summed or segmented dispersion relationship exhibiting a region conforming to Gaussian behavior must be investigated carefully to determine whether it is created by background subtraction or a separate underlying process. Equations 3, 7, and 10 can be used to better understand the impact of offsets or instrument pre-processing in a data set. In our data, the background and analyte responses conformed to a single set of dispersion and process parameters. By avoiding the subtraction step which produces a response of the form of Equation 3, a generalized difference detection limit (DDL) can be directly defined on the dispersion model.

For all systems for which a dispersion model can be measured, a DDL can be defined which is the minimum distance between two mean values $\bar{x}_1$ and $\bar{x}_2$ required to meet a particular

statistical test at a specified confidence level. The strength of this definition is that it avoids the calibration steps which may distort an otherwise simple relationship and applies to any pair of mean values at any position along a dispersion model not just the region used for conventional detection limits.

Both instruments gave a calibrated linear response over approximately 6 orders of magnitude. Calibrating the data set in concentration units produced an uncertainty function resembling a bent stick or root quadratic models. The position of the bend varied between the two instruments and the two emission lines investigated. It also revealed a zone between 1 and 300 ppm where the reproducibility of the two instruments was identical. Across all the presentations, no fluctuation scaling exponent ever matched the Horwitz exponent ($\alpha = 0.85$) and did not generally match the linear portion of the characteristic function with an exponent significantly above 1. This inter-laboratory study demonstrates how fluctuation scaling studies provide a framework for robustly comparing instrumentation across multiple laboratories.

Acknowledgements

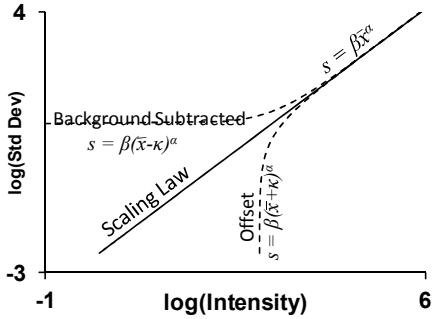
The authors wish to thank Nigel Mould (NTU) and Diane Spencer (Wolverhampton) for technical support for these measurements.

References

(1) Zitter, H.; God, C. *Fresenius' Journal of Analytical Chemistry* **1971**, *255*, 1-9.
(2) Thompson, M. *J. AOAC Int.* **2012**, *95*, 1803-1806.
(3) Thompson, M. *TrAC, Trends Anal. Chem.* **2011**, *30*, 1168-1175.
(4) Thompson, M. *Analyst* **2000**, *125*, 385-386.
(5) Thompson, M. *Analyst* **1999**, *124*, 991-991.
(6) Thompson, M.; Howarth, R. J. *Journal of Geochemical Exploration* **1978**, *9*, 23-30.
(7) Horwitz, W. *Anal. Chem.* **1982**, *54*, 67A-76A.
(8) Albert, R.; Horwitz, W. *Anal. Chem.* **1997**, *69*, 789-790.
(9) Horwitz, W.; Britton, P.; Chirtel, S. J. *J. AOAC Int.* **1997**, *81*, 1257-1265.

(10) Rocke, D. M.; Lorenzato, S. *Technometrics* **1995**, *37*, 176-184.
(11) Voigtman, E. In *Limits of Detection in Chemical Analysis*, Vitha, M. F., Ed.; Wiley, 2017, pp 199-214.
(12) Hanley, Q. S. *Anal. Chem.* **2016**, *88*, 12036-12042.
(13) Kendal, W. S. *Proceeding of the Royal Society A* **2017**, *473*, 20160586.
(14) Kendal, W. S.; Jørgensen, B. *Computation* **2015**, *3*, 528-540.
(15) Kendal, W. S.; Jørgensen, B. *Phys. Rev. E* **2011**, *83*, 066115.
(16) Horwitz, W.; Albert, R. *J. AOAC Int.* **2006**, *89*, 1095-1109.
(17) Eisler, Z.; Bartos, I.; Kertész, J. *Adv. Phys.* **2008**, *57*, 89-142.
(18) Cohen, J. E.; Xu, M.; Schuster, W. S. *P. Roy. Soc. Lond. B. Bio.* **2013**, *109*, 15829–15834.
(19) Taylor, L. *Nature* **1961**, *189*, 732-735.
(20) El-Shaarawi, A. H.; Zhu, R.; Joe, H. *Environmetrics* **2011**, *22*, 152-164.
(21) Jørgensen, B.; Kokonendji, C. C. *AStA Adv. Stat. Anal.* **2016**, *100*, 43-78.
(22) Jørgensen, B. *The Theory of Dispersion Models*; Chapman & Hall, 1997; Vol. 76, p 237.
(23) Prudnikov, E. D.; Elgersma, J. W.; Smit, H. C. *Journal of analytical atomic spectrometry* **1994**, *9*, 619-622.
(24) Mortara, L.; Fowler, A. *Proc. SPIE, 0290* **1981**, 28-33.
(25) Mackay, C. D. *Annual Review of Astronomy and Astrophysics* **1986**, *24*, 255-283.
(26) Prabhu, T.; Mayya, Y.; Anupama, G. *Journal of astrophysics and astronomy* **1992**, *13*, 129-144.
(27) Robbins, M. S.; Hadwen, B. J. *IEEE Transactions on Electron Devices* **2003**, *50*, 1227-1232.
(28) Hanley, Q.; Khatun, S.; Yosef, A.; Dyer, R.-M. *PLoS ONE* **2014**, *9*, e109004.
(29) Anscombe, F. J.; Glynn, W. J. *Biometrika* **1983**, *70*, 227-234.
(30) Voigtman, E. *Limits of Detection in Chemical Analysis*; Wiley, 2017; Vol. 185, p 199-214.
(31) Cox, N. J. *Stata Journal* **2010**, *10*, 482.
(32) Wilkins, J. E. *Ann. Math. Stat.* **1944**, *15*, 333-335.
(33) Dalén, J. *Statistics & Probability Letters* **1987**, *5*, 329-331.
(34) Medina, J. M.; Díaz, J. A. *Progress in Organic Coatings* **2017**, *104*, 118-124.

For TOC Only


4
-3
-1
6
log(Std Dev)
log(Intensity)
Background Subtracted
$s = \beta(\bar{x}-\kappa)^{\alpha}$
Scaling Law
Offset
$s = \beta(\bar{x}+\kappa)^{\alpha}$
$s = \beta\bar{x}^{\alpha}$